11th World Congress on Structural and Multidisciplinary Optimisation
07th -12th, June 2015, Sydney Australia

Experience with Several Multi-fidelity Surrogate Frameworks

Chanyoung Park[1], Raphael T. Haftka[2] and Nam-Ho Kim[2]

[1] University of Florida, Gainesville, FL, USA, cy.park@ufl.edu
[2] University of Florida, Gainesville, FL, USA, haftka@ufl.edu
[3] University of Florida, Gainesville, FL, USA, nkim@ufl.edu

1. Abstract
In this paper, multi-fidelity surrogate (MFS) frameworks are investigated with the aid of an algebraic example for 100 different designs of experiments (DOEs). These include three Bayesian frameworks using 1) a model discrepancy function, 2) low fidelity model calibration and 3) a comprehensive approach. Two simple frameworks using 1) a discrepancy function and 2) low fidelity model calibration which are counterparts of the Bayesian frameworks 1) and 2) are also investigated. Their computational cost saving and accuracy improvement over a single fidelity surrogate model are investigated as a function of the ratio of the sampling costs of low and high fidelity simulations. The maximum cost saving was 85% and the maximum accuracy improvement was 40% when the number of low fidelity samples is about ten times larger than that of high fidelity samples for various computational costs. We found that the DOE can substantially change the relative standing of the different frameworks. Therefore, an important question is how to determine which model works best for a specific problem and DOE. The cross validation error appears to be a reasonable candidate for estimating which MFS models would perform poorly for a specific problem.
2. Keywords: Multi-fidelity surrogate framework, Comparison study, Discrepancy function, Calibration, Bayesian

3. Introduction
Surrogate models have been used as a cheap approximations, which can be built with several dozen data points. However, for sophisticated high fidelity models, the cost for obtaining sufficient data for achieving reasonable accuracy of a surrogate is high. Multi-fidelity surrogate (MFS) models have been developed to compensate for inadequate expensive high fidelity data with cheap low fidelity data by modelling the connection between two models. There are Gaussian process (GP) based Bayesian MFS frameworks and simple frameworks using the Kriging surrogate model which is a regular surrogate model based on GP, as well as other surrogates. However, the performance of different MFS frameworks with different complexities has been rarely compared.

Building an MFS using a model discrepancy function is a popular framework. An MFS is built by combining a low-fidelity surrogate based on low fidelity data set and a discrepancy surrogate based on low and high fidelity data sets. This multi-fidelity framework has been used in design optimization to alleviate computational burden. For example, non-Bayesian MFS models using linear regression were used by combining coarse and fine finite element models for aircraft structural optimization [1,3]. The same approach was used to combine aerodynamic prediction from cheap linear theory and expensive Euler solutions for aircraft aerodynamic optimization [2]. A Bayesian MFS model based on GP was later introduced by Kennedy and O'Hagan (2000). The Bayesian model allows to incorporate prior information [4,5]. Co-Kriging [9,10] provides an equivalent result for the Bayesian formulation with a non-informative prior and has good computational characteristics [6,7,8].

Model calibration is another MFS model based on tuning model parameters of a low fidelity model. A straight forward simple framework is to find parameters that minimize discrepancy between a calibrated low fidelity model and high fidelity data [14]. GP based Bayesian calibration frameworks were also introduced [11,12,13]. The Bayesian frameworks find best calibration parameters that is most statistically consistent with available information [11,15]. A general Bayesian MFS model that uses both calibration and a discrepancy function was proposed by Kennedy and O'Hagan (2001) offering greater flexibility.

The objectives of this paper are: (1) introduce characteristics of multi-fidelity surrogate frameworks, (2) investigate the performance of those surrogates in terms of accuracy (3) investigate the performance of the cross validation error as a surrogate performance estimator. This paper is organized as follows. Section 2 presents multi-fidelity surrogate models. Section 3 describes the methodology of the investigation. Section 4 presents results and discussion.

2. Multi-fidelity surrogate models with different frameworks
In this paper, we discuss three commonly used Bayesian MFS frameworks: using (1) a model discrepancy function, (2) a low fidelity model parameter calibration and (3) a combined approach. Two simpler MFS frameworks are

also discussed, which are simplified versions of the Bayesian MFS frameworks using the approach (1) and (2). The characteristics of frameworks and the differences between Bayesian and simple frameworks will be described in the following sections.

2.1 Simple MFS framework using a model discrepancy function

This framework provides a convenient way of fitting an MFS with regular surrogate models. In this framework the MFS is described with two surrogates $\hat{y}_{low}(\mathbf{x})$ and $\hat{\delta}(\mathbf{x})$ which represent a low fidelity function and a discrepancy function, respectively, as

$$\hat{y}_H(\mathbf{x}) = \rho \hat{y}_L(\mathbf{x}) + \hat{\delta}(\mathbf{x}) \tag{1}$$

where ρ is a regression scalar minimizing $\left(\rho \hat{y}_L(\mathbf{x}_H) - \mathbf{y}_H\right)^T \left(\rho \hat{y}_L(\mathbf{x}_H) - \mathbf{y}_H\right)$

Figure 1 illustrates an example with a high fidelity function $y_{HT}(x) = (6x-2)^2 \sin(12x-4)$ and a low fidelity function $y_{LT}(x) = (5.5x-2.5)^2 \sin(12x-4)$ and the corresponding data sets with 5 and 24 samples, respectively. The high fidelity sampling point set is a subset of the low fidelity sampling point set.

Figure 2 (a) and (b) show fits by combining the simple MFS framework and the Kriging surrogate model that a low fidelity Kriging surrogate is fitted with the low fidelity data set and a Kriging surrogate of the discrepancy function is fitted using the difference at the common sampling points. To understand the effect of the regression scalar ρ, we fitted the surrogates with and without a condition ρ=1.

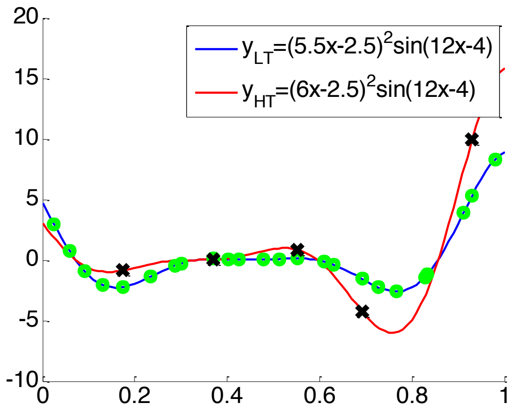


Figure 1: Low fidelity function and high fidelity function and low fidelity data (green circles) and high fidelity data (black crosses)

2.2 Bayesian MFS framework using a model discrepancy function

In the previous section, the simple framework was described. When the framework combined with a Kriging surrogate, it is a special case of the full Bayesian framework which was introduced [4,5]. The main differences between the simple framework and this Bayesian framework are: 1) the simple framework uses the low fidelity data set while the Bayesian framework uses high and low fidelity data sets to update the low fidelity response prior model and 2) whole low fidelity data is used to get the discrepancy function whereas the simplified framework use low fidelity data at the common points.

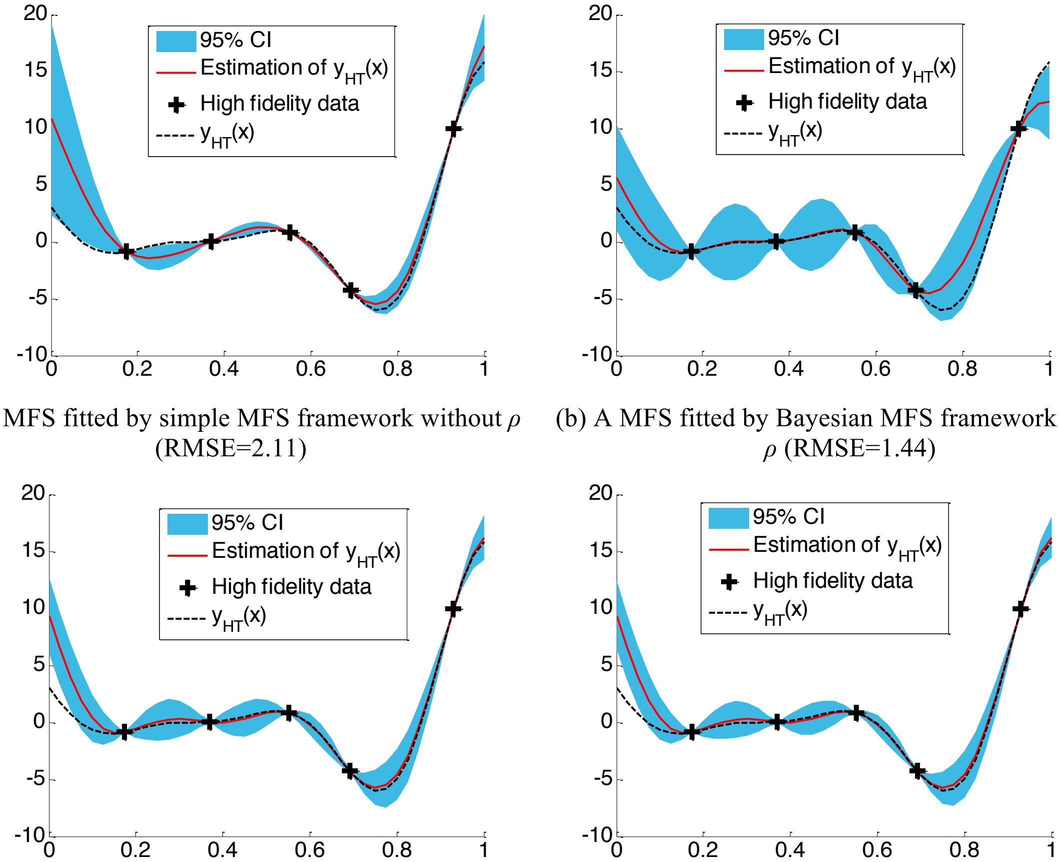
Figure 2 (c) and (d) present fits using the Bayesian MFS framework with and without the condition ρ=1 and the corresponding prediction uncertainties for 95% confidence. From a comparison between (a) to (d), the simple framework can provide a fit as good as Bayesian framework and the effect of the regression scalar is more important than applying the Bayesian framework for this example.

2.3 Simple MFS framework with model calibration

The previous two frameworks try to capture the discrepancy between high and low fidelity responses using a model discrepancy function. An widely used alternative framework is to tune parameters of a low fidelity model which is also known as model calibration. A simple calibration based framework is to build a surrogate of the low fidelity response $\hat{y}_L(\mathbf{x},\mathbf{q})$ which is a function of the input variables $\boldsymbol{x}$ and the model parameters $\boldsymbol{q}$ to define the response as a function of model parameters. Then we can find optimal $\mathbf{q} = \boldsymbol{\theta}$ minimizing the discrepancy between the high fidelity data and the low fidelity prediction.

$$\hat{y}_H(\mathbf{x}) = \hat{y}_L(\mathbf{x},\boldsymbol{\theta}) \tag{2}$$

Figure 3 (b) shows a fit using the Kriging surrogate model and the simple framework with the previous 1-D function example with the selection of calibration parameters as $y_{LT}\left(x,\{\theta_1,\theta_2\}\right) = (\theta_1 x - \theta_2)^2 \sin(12x-4)$. Since the low fidelity function has the same function form with the high fidelity function, the calibrated parameters should be θ_1=6, θ_2=-2, and Fig. 3 (a) shows the sampling points to fit the surrogate.



(a) A MFS fitted by simple MFS framework without ρ (RMSE=2.11)

(b) A MFS fitted by Bayesian MFS framework without ρ (RMSE=1.44)

(c) A MFS fitted by simple MFS framework with ρ (RMSE=1.40 for ρ=1.72)

(d) A MFS fitted by Bayesian MFS framework with ρ (RMSE=1.38 for ρ=1.70)

Figure 2: Multi-fidelity surrogates using a model discrepancy function

2.4 Bayesian MFS framework with model calibration

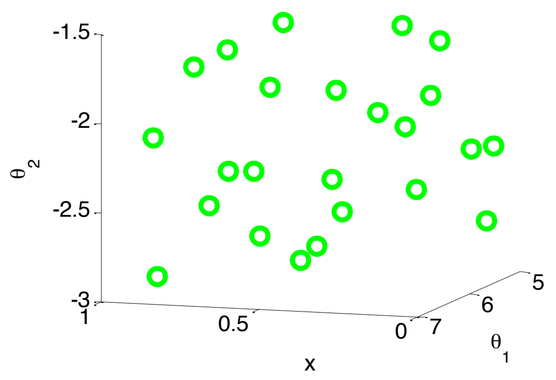
The Bayesian MFS framework makes inference about $y_{HT}(\mathbf{x})$ by updating a high fidelity prior model based on low and high fidelity data sets [11,15]. A big difference from the previous simple framework is that the Bayesian framework captures the model inadequacy of a low fidelity function as well as calibrating the low fidelity model while the previous simple framework cannot capture the inadequacy. As Fig. 3 (c) shows, predictions at high fidelity data points perfectly match data even with the model inadequacy of the low fidelity model and the prediction uncertainties at those points are zero. In Fig. 3 (d), the comprehensive Bayesian framework calibrates parameters well but its fit is worse than the Bayesian framework without a model discrepancy function.

2.5 Comprehensive Bayesian MFS framework

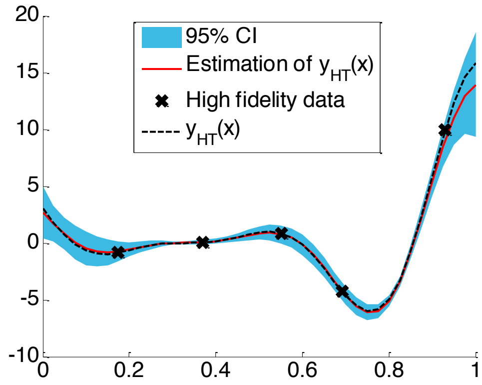
A comprehensive MFS is a most flexible model that the model discrepancy between low and high fidelity model responses is captured by tuning low fidelity model parameters and a model discrepancy function [12,13].

2.6 Strategies for design of experiments

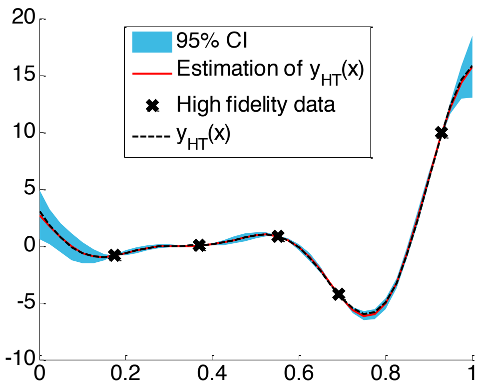
Building multi-fidelity surrogates requires a new sampling strategy since MFSs need low and high fidelity data sets and it is natural that the low fidelity data set is a super set of high fidelity data set to see the model discrepancy. A sampling strategy, which is called nested neighborhood design, is to generate low fidelity sampling points using LHS and to select high fidelity sampling points from the generated low fidelity sampling points having optimal coverage [7]. All the examples in this paper use the nested neighborhood design.



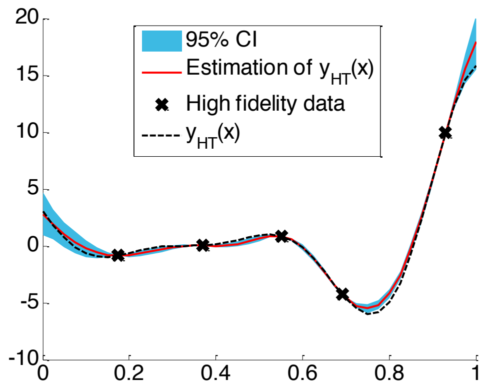
(a) 24 sampling points on $\{x, \theta_1, \theta_2\}$ space



(b) Simple MFS fit with model calibration (RMSE=0.47 for θ_1=6, θ_2=-2, ρ=1)



(c) Bayesian MFS fit with model calibration (RMSE=0.10 for θ_1=6, θ_2=-2, ρ=1)



(d) Comprehensive Bayesian MFS fit (RMSE=0.45 for θ_1=6, θ_2=-2, ρ=1)

Figure 3: Multi-fidelity surrogates based on model calibration

3. Measurements of a surrogate model

In this section, we describe numerical experiments for assessing the robustness of prediction of the MFS frameworks. Since the performance of surrogates varies for different problems and design of experiments (DOE), cross validation error has been used as a measure to rank for surrogate models. We investigate whether the cross validation error can be employed to detect the worst MFS frameworks for a specific problem and specific DOE.

3.1 Assessing accuracy using root mean square error

MFS surrogates are fitted to function values at n points which are generated by the sampling strategy. We can measure the accuracy of an MFS using the RMSE which is the square root of square error integrated over the sampling domain. We use Monte Carlo integration at a large number of n_{test} test points as

$$\text{RMSE} = \sqrt{\frac{1}{n_{test}}\sum_{i=1}^{n_{test}}\left(y_i - \hat{y}_i\right)^2} \qquad \text{for } i = 1, \ldots n_{test} \tag{3}$$

where $\hat{y}_i$ and y_i are a prediction and a true function value at the i^{th} test point.

4. Numerical examples

We compare the previously described frameworks with the Hartmann 6 function example. Their accuracy was statistically assessed since their performances vary for different DOEs. 100 different DOEs were generated using nested neighborhood sampling and their accuracies were measured with the median RMSE of 100 RMSEs.

The computational cost saving for achieving a certain accuracy is main thrust of applying MFS frameworks for fitting a surrogate. In this section, we examine the described frameworks in that perspective. There are other factors that determine the efficiency of an MFS framework: 1) the ratio of high fidelity sample size to low fidelity sample size, 2) the ratio of the high fidelity data point evaluation cost to the low fidelity data point cost and 3) the total computational budget. We also consider those three factors.

Table 1 shows cases for different factors that we discuss in this paper. The computational budget 56H means we have computational budget equivalent to evaluating 56 high fidelity samples. Here we use the high fidelity sample evaluation cost as a reference cost since typically we need to fit an MFS for a given high fidelity simulation that we have no choice but to choose a low fidelity simulation. Cost ratio 4 means 4 low fidelity samples can be evaluated with the budget for evaluating a single high fidelity sample. The cases show combinations of low and high fidelity samples for given total budget and cost ratio. For example, 36/80 of the budget of 56H and cost ratio 4 denotes a case with 36 high fidelity samples and 80 low fidelity samples for fitting an MFS. There are different cases for given computational budget and sample cost ratio. For example, the computational budget of 56H and a ratio of 4, leads to 36H+80/4H=56H.

Table 1: Computational budget and a data point evaluation ratio

Computational budget	LF sample cost ratio	Sample size ratio
56H	4	36/80, 26/120, 16/160, 6/200
	10	49/70, 46/100, 42/140, 35/210, 28/280, 21/350, 14/420, 7/490
	30	48/240, 46/300, 44/360, 42/420, 40/480, 38/540, 28/840, 18/1140
28H	4	18/40, 13/60, 8/80, 3/100
	10	22/60, 19/90, 16/120, 13/150, 10/180, 7/210, 4/240
	30	24/120, 22/180, 20/240, 18/300, 16/360, 14/420, 10/540, 6/660, 4/720

We use the Hartmann 6 function [16] over [0.1,1] as a high fidelity function and an approximated function of the Hartmann function as a low fidelity function. The approximated function uses a different alpha $\boldsymbol{\alpha}_{approx} = \{0.5 \quad 0.5 \quad 2.0 \quad 4.0\}^T$ and an approximated exponential function which is expressed as

$$f_{approx}^{\exp}(x) = \left(\exp\left(\frac{-4}{9}\right) + \exp\left(\frac{-4}{9}\right)\frac{(x+4)}{9} \right)^9 \tag{4}$$

We generated 100 different DOEs using the nested neighborhood sampling to check the robustness of MFS frameworks. RMSE of each DOE was calculated for each framework based on the same 10,000 test points generated by LHS. The medians of 100 RMSEs were obtained for each framework and compared to one another. Table 2 shows the effect of cost saving for the best framework for different cases. The RMSEs of single fidelity surrogates show RMSE with the corresponding computational budget so that RMSE for the high fidelity fit remains the same for all cases. The cost saving is the computational cost saving by using the best MFS framework for the cost of the high fidelity fit which achieves the same level of RMSE with the corresponding MFS framework.

Table 2: The effect of cost saving for the best framework for the best sample size ratio

Computational budget	LF sample cost ratio	RMSEs of single fidelity surrogates	Best RMSE	Best sample size ratio	Cost saving
56H	4	$RMSE_L$=0.123 $RMSE_H$=0.132	0.095 (Bayesian disc.)	[26/120, 16/160]	50%
	10	$RMSE_L$=0.116 $RMSE_H$=0.132	0.072 (Bayesian disc.)	[28/280, 21/350]	75%
	30	$RMSE_L$=0.113 $RMSE_H$=0.132	0.06 (Bayesian disc.)	[44/360, 40/480]	83%
28H	4	$RMSE_L$=0.136 $RMSE_H$=0.166	0.125 (Bayesian disc.)	80/8	58%
	10	$RMSE_L$=0.122 $RMSE_H$=0.166	0.1 (Bayesian disc.)	[13/150, 7/210]	75%
	30	$RMSE_L$=0.114 $RMSE_H$=0.166	0.08 (Bayesian disc.)	[20/240, 14/420]	85%

5. Concluding Remarks

In this paper, we present a comparison study of five MFS building frameworks by combining low and high fidelity data sets using the 6D Hartmann 6 function: 1) a simple framework based on a discrepancy function approach using the Kriging surrogate model, 2) a Bayesian framework based on a discrepancy function, 3) a simple framework based on model calibration using the Kriging surrogate model, 4) a Bayesian framework based on model calibration and 5) a comprehensive Bayesian framework.

We found that the MFS frameworks become useful as the cost of a low fidelity data point becomes cheaper than the cost of a high fidelity data point and the MFS frameworks become most beneficial for saving the computational cost. Based on the example, computational cost can be saved by 85% for the same accuracy with the single fidelity surrogate while the maximum accuracy improvement over the single fidelity surrogate is 60% improvement in terms of RMSE.

For the discrepancy function based frameworks, an interesting observation was that the simple framework could perform as well as the Bayesian framework and the use of the regression scalar ρ could be important. For the calibration based frameworks, the framework without a discrepancy function outperformed the comprehensive framework and the effect of the regression scalar ρ was not noticeable. In terms of accuracy, the discrepancy function based frameworks showed good performance generally. The calibration frameworks could show reliable performance with a few high fidelity samples whereas the performance of the discrepancy function based frameworks abruptly decreased for a few high fidelity samples but there were factors that might affect the results such as calibration parameter selection and calibration parameter bounds which were not seriously considered in this paper. There were optimal ratios maximizing the accuracy but the effect was not significant when sufficient number of low fidelity data points were obtained.

6. Acknowledgements
This work is supported by the U.S. Department of Energy, National Nuclear Security Administration, Advanced Simulation and Computing Program, as a Cooperative Agreement under the Predictive Science Academic Alliance Program, under Contract No. DE-NA0002378.

7. References

[1] Balabanov, V., Haftka, R. T., Grossman, B., Mason, W. H., & Watson, L. T. (1998, September). Multifidelity response surface model for HSCT wing bending material weight. In *Proceedings of 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization* (pp. 778-788).

[2] Knill, D. L., Giunta, A. A., Baker, C. A., Grossman, B., Mason, W. H., Haftka, R. T., & Watson, L. T. (1999). Response surface models combining linear and Euler aerodynamics for supersonic transport design. *Journal of Aircraft*, *36*(1), 75-86.

[3] Mason, B. H., Haftka, R. T., Johnson, E. R., & Farley, G. L. (1998). Variable complexity design of composite fuselage frames by response surface techniques. *Thin-walled structures*, *32*(4), 235-261.

[4] Kennedy, M. C., & O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, *87*(1), 1-13.

[5] Qian, P. Z., & Wu, C. J. (2008). Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. Technometrics, 50(2), 192-204.

[6] Forrester, A. I., Sóbester, A., & Keane, A. J. (2007). Multi-fidelity optimization via surrogate modelling. Proceedings of the royal society A: mathematical, physical and engineering science, 463(2088), 3251-3269.

[7] Le Gratiet, L. (2013). Multi-fidelity Gaussian process regression for computer experiments (Doctoral dissertation, Université Paris-Diderot-Paris VII).

[8] Kuya, Y., Takeda, K., Zhang, X., & J. Forrester, A. I. (2011). Multifidelity surrogate modeling of experimental and computational aerodynamic data sets. *AIAA journal*, *49*(2), 289-298.

[9] Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical science*, 409-423.

[10] Lophaven, S. N., Nielsen, H. B., & Søndergaard, J. (2002). *DACE-A Matlab Kriging toolbox, version 2.0*.

[11] McFarland, J., Mahadevan, S., Romero, V., & Swiler, L. (2008). Calibration and uncertainty analysis for computer simulations with multivariate output. AIAA journal, 46(5), 1253-1265.

[12] Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(3), 425-464.

[13] Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., & Ryne, R. D. (2004). Combining field data and computer simulations for calibration and prediction.*SIAM Journal on Scientific Computing*, *26*(2), 448-466.

[14] Zheng, L., Hedrick, T. L., & Mittal, R. (2013). A multi-fidelity modelling approach for evaluation and optimization of wing stroke aerodynamics in flapping flight. *Journal of Fluid Mechanics*, *721*, 118-154.

[15] Prudencio, E. E., & Schulz, K. W. (2012, January). The parallel C++ statistical library 'QUESO': Quantification of Uncertainty for Estimation, Simulation and Optimization. In *Euro-Par 2011: Parallel Processing Workshops* (pp. 398-407). Springer Berlin Heidelberg.

[16] http://www.sfu.ca/~ssurjano/hart6.html